SPi	Journal Code:				Article ID				Dispatch: 1.0.1	CE: Belinda Pinili
	P	S	T		1	7	7	4	No. of Pages: 9	ME:

MAIN PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1774

Published online in Wiley Online Library

A novel approach to estimation of the time to biomarker threshold: applications to HIV

Tarylee Reddy,[a,b]* Geert Molenberghs,[b,c] Edmund Njeru Njagi,[b] and Marc Aerts[b]

In longitudinal studies of biomarkers, an outcome of interest is the time at which a biomarker reaches a particular threshold. The CD4 count is a widely used marker of human immunodeficiency virus progression. Because of the inherent variability of this marker, a single CD4 count below a relevant threshold should be interpreted with caution. Several studies have applied persistence criteria, designating the outcome as the time to the occurrence of two consecutive measurements less than the threshold. In this paper, we propose a method to estimate the time to attainment of two consecutive CD4 counts less than a meaningful threshold, which takes into account the patient-specific trajectory and measurement error. An expression for the expected time to threshold is presented, which is a function of the fixed effects, random effects and residual variance. We present an application to human immunodeficiency virus-positive individuals from a seroprevalent cohort in Durban, South Africa. Two thresholds are examined, and 95% bootstrap confidence intervals are presented for the estimated time to threshold. Sensitivity analysis revealed that results are robust to truncation of the series and variation in the number of visits considered for most patients. Caution should be exercised when interpreting the estimated times for patients who exhibit very slow rates of decline and patients who have less than three measurements. We also discuss the relevance of the methodology to the study of other diseases and present such applications. We demonstrate that the method proposed is computationally efficient and offers more flexibility than existing frameworks.

Keywords: threshold; prediction; HIV progression; CD4 count; persistence criteria; seroprevalent cohort

1. INTRODUCTION

Biomarkers are widely used in the screening, diagnosis and monitoring of several diseases. In screening programmes, cohort studies and clinical trials where interest lies in a repeatedly measured biomarker, the outcome of interest may be the time at which a biomarker reaches a particular threshold. Biomarkers for clinical events are particularly useful in the study of the human immunodeficiency virus (HIV) progression due to the long natural history of the disease. CD4 cells, which are lymphocyte cells responsible for immune response to infections, are the primary target of the virus and are cited as a relevant predictor of acquired immunodeficiency syndrome (AIDS) related illness and death [1]. A comprehensive review of CD4 count as a surrogate marker for AIDS defining illnesses and death can be found in [2]. Another relevant biomarker is viral load (VL), which quantifies the level of HIV in blood. VL is used as a surrogate marker for treatment efficacy, AIDS-defining illnesses and death [3]. The World Health Organization (WHO) guidelines recommend treatment for all HIV-positive patients, but emphasize treatment as a priority for patients with CD4 counts less than 350 cells/mm^3 or WHO stage 3 or 4 symptoms [4]. Hence, in the absence of clinical manifestations, the urgency of treatment initiation is based on a single CD4 count measurement less than 350 cells/mm^3. The time to reach a relevant CD4 count threshold is also used as an endpoint in HIV/AIDS clinical trials as a marker of treatment success and in cohort studies where interest is in examining HIV progression.

A study of variability of CD4 count in patients enrolled in the ACTG trial revealed that measurements taken 8 weeks apart differed by 20% [5]. [6] and [7] attribute this variability in CD4 counts to several factors including diurnal variation, measurement error, psychological and physical stress, diet and the menstrual cycle. Owing to the inherent variability of this marker, a single CD4 count below a relevant threshold should be interpreted with caution.

1.1. Current approaches to time to threshold modelling

1.1.1. Standard approaches. The time to CD4 threshold has been analysed as an outcome in several recent studies. The convention in such studies is to first extract the time of the event, which is analysed in a second stage within the survival analysis framework. [8] analysed the time to first CD4 count less than 350 cells/mm^3 as the primary outcome in their study, which compared the rate of HIV progression pre and post the introduction of antiretroviral therapy (ART). In a recent application, [9] examined the effect of the rs12252-C genotype on HIV progression, which was defined

[a] *Biostatistics Unit, Medical Research Council, Durban, South Africa*

[b] *I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium*

[c] *I-BioStat, KU Leuven, Leuven, Belgium*

**Correspondence to: Tarylee Reddy, Biostatistics Unit, Medical Research Council, Durban, South Africa.*
E-mail: tarylee.reddy@mrc.ac.za

as the time of the first CD4 count less than 350 cells/mm^3. [10] studied a cohort of recently infected individuals where the effect of HIV subtype on HIV progression was examined. In this study, immunologic progression was defined as time from seroconversion to the first of two consecutive CD4 cell counts less than or equal to 350 cells/mm^3. Imposing persistence criteria such as a 'two consecutive' rule is an improvement on basing clinical decisions on a single CD4 count, but can be unreliable when the time between visits is large. In addition, patients who enter the study with a CD4 count already below the relevant threshold, typically are excluded in modelling time to threshold. Removing these patients from the analysis results in left truncation and biased inferences.

1.1.2. General model-based approaches. A criticism of the standard approach discussed earlier is that it ignores the inherent subject-specific CD4 count trajectory and assumes that the event times are observed without error. Model-based approaches to the time to threshold of a biomarker have recently emerged. One such approach is inverse estimation or calibration. [11] examined the time for subjects enrolled in the 'Multicentre aneurysm screening study' to reach an aneurysm diameter of 55 mm. The researchers used the method of inverse estimation in linear and quadratic subject-specific curves from a Bayesian hierarchical model. The estimated times to threshold were found to be too imprecise to be used in practice. In cases where the interest lies in modelling the time to threshold of an ordinal variable, continuous time Markov models have proven to be useful. This involves forming a distinct set of states and computing the mean first hitting time to a particular state. [12] added to this methodology by studying the first hitting time to a state followed by a fixed duration of stay in the state. This was applied to a study on multiple sclerosis where sustained progression based on a disability scale was of interest. When the outcome of interest is a continuous biomarker, the construction of discrete states is somewhat arbitrary. Furthermore, due to the high degree of variability of the biomarker, reverse transitions and transitions that skip intermediate states are often observed [13].

1.1.3. Contribution and organization of this paper. In this paper, we propose an approach to time to threshold modelling that involves two stages. In the first stage, a linear mixed model is fitted to the longitudinal measurements, resulting in patient-specific predicted values that are a function of the fixed effects and empirical Bayes estimates. In the second stage, the probability of experiencing two consecutive measurements less than a relevant threshold k at each time point is computed and substituted into the expression for the expected time to threshold. The methodology underlying our proposed method is presented in Section 2 and the Appendix. This approach, which was motivated by time to threshold modelling in the HIV setting, is applied in Section 3 to data emanating from a cohort of HIV-positive individuals in South Africa. For ease of presentation, we draw attention to the estimated time to threshold for four selected patients, each of whom exhibits a different CD4 count evolution. We assess the sensitivity of results to changes in the time points considered and present these results in Section 3.4. In addition to the HIV application, in Section 4, we discuss the relevance of the methodology to the study of other diseases and present three specific applications. In Section 5, we discuss the performance and flexibility of our proposed approach and present areas of further work.

2. METHODOLOGY

2.1. Expected time to attain threshold

We propose a method to estimate the time to attain two consecutive CD4 counts less than or equal to the relevant threshold k. For readability, we will refer to the attainment of two consecutive CD4 counts less than the threshold, as the event of interest. In the approach we propose, we consider the individual 'at risk' for the event both prior to and post enrolment. Letting Y_{ij} denote the CD4 count observed on individual i at time point j, where $j = 1$ corresponds to an occasion at which one starts considering the individual as possibly seroconverting, the time to event T_i can be expressed as

$$T_i = \min\{j \geq 2 : Y_{ij-1} \leq k, Y_{ij} \leq k\}. \tag{1}$$

It follows that the expected time for individual i to attain two consecutive CD4 counts less than the threshold k can be expressed as follows:

$$\begin{aligned} E(T_i) &= t_{i2}P(Y_{i1} \leq k, Y_{i2} \leq k) + t_{i3}P(Y_{i1} > k, Y_{i2} \leq k, Y_{i3} \leq k) \\ &\quad + t_{i4}\{P(Y_{i1} > k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k) \\ &\quad + P(Y_{i1} \leq k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k)\} \\ &\quad + \ldots \\ &= \sum_{j=2}^{\infty} t_{ij}S_{ij}, \end{aligned} \tag{2}$$

where t_{ij} represents the time corresponding to the jth visit for individual i, and S_{ij} denotes the probability of individual i experiencing the event, or 'stopping', at t_{ij}. In practice, the infinite series may be truncated at a time point considered relevant to the specific application at hand. Possible options for the time at which the series is truncated are the expected lifetime of an individual, or the end of the incubation period of a particular disease. We specify a linear mixed model, which satisfies

$$\begin{aligned} \boldsymbol{y}_i &= X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{b}_i &\sim N(\boldsymbol{0}, D), \\ \boldsymbol{\varepsilon}_i &\sim N(\boldsymbol{0}, \Sigma_i), \end{aligned} \tag{3}$$

where $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\varepsilon}_i, \ldots, \boldsymbol{\varepsilon}_N$ are independent. $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ represent the fixed and random effects, respectively [14]. It follows that

$$\mathbf{y}_i|\mathbf{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i).$$

As presented in (2), S_{ij} is the sum of several joint probabilities, each of which represents a distinct combination of the values of Y_{ij} that may yield the event at time point j. Assuming conditional independence in (3) such that $\Sigma_i = \sigma^2 I_{n_i}$, the joint probabilities that form S_{ij} reduce to the product of the individual probabilities. Hence, S_{ij} may be simplified as follows:

$$\begin{aligned} S_{ij}(X_i, Z_i, \mathbf{b}_i, \boldsymbol{\beta}) &= C_{ij-3}P(Y_{ij-2} > k)P(Y_{ij-1} \leq k)P(Y_{ij} \leq k) \\ &= C_{ij-3}[1 - \tilde{\Phi}_{ij-2}(k)][\tilde{\Phi}_{ij-1}(k)][\tilde{\Phi}_{ij}(k)], \end{aligned}$$

where C_{ij-3} denotes the 'continuation probability' at time t_{ij-3} and $\tilde{\Phi}_{ij}(k)$ is a cumulative normal distribution with mean $\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i$ and variance σ^2. It follows that $\tilde{\Phi}_{ij}(k)$ can be expressed as a simple function of the standard univariate normal distribution:

$$\tilde{\Phi}_{ij}(k) = \Phi\left(\frac{k - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{b}_i}{\sigma}\right). \quad (4)$$

Therefore, the estimated individual probability $\tilde{\Phi}_{ij}(k)$ is a function of the fixed effects estimates, empirical Bayes estimates and measurement error. For a model with a strictly decreasing trend in t_{ij}, at a fixed threshold k, one would expect the probability in (4) to decrease with increasing j. The assumption of conditional independence in (3) and the recursive relationship of continuation probabilities simplify the computation of $E(T_i)$, but may be extended to accommodate correlated residuals. Admittedly, more complex multivariate normal probability manipulation will then be necessary.

The continuation probability C_{ij} can also be interpreted in the survival analysis framework as the probability of individual i being at risk for the event after time t_{ij}. That is, the probability that individual i has not experienced two consecutive low CD4 counts at, or prior to time point j. It should be evident that as j increases the computation of C_{ij} will become increasingly complex due to the number of combinations considered. Careful examination of the pattern governing the number of combinations that result in continuation at each time point revealed the recursive relationship

$$C_{ij} = C_{j-2}[1 - \Phi_{ij-1}(k)][\Phi_{ij}(k)] + C_{j-1}[1 - \Phi_{ij}(k)]. \quad (5)$$

Further details regarding the proof of (5) and its computation can be found in the Appendix.

2.2. Estimation and inference

It follows from Section 2.1 that $E(T_i)$ is a function of the parameters $\boldsymbol{\beta}$, $\boldsymbol{b}_i$ and σ. Hence $\widehat{T}_i$, the estimate of $E(T_i)$, can be computed by substituting each unknown parameter by its corresponding estimate. Further details on inference for fixed effects and empirical Bayes prediction of the random effects in a linear mixed model can be found in [14]. In principle, the Delta method may be used to compute standard errors and 95% confidence intervals for $\widehat{T}_i$. However, the bootstrap offers a more feasible alternative. We propose a conditional version of the non-parametric bootstrap to compute 95% confidence intervals for $\widehat{T}_i$ as follows:

Step 1. Individual i is removed from the full dataset resulting in $N - 1$ cases
Step 2. Sample $N - 1$ subjects with replacement from the dataset in step 1
Step 3. Append the data of individual i to the bootstrap sample
Step 4. Compute $\widehat{T}_i$

This process is repeated 1000 times.

3. APPLICATION: REPEATED CD4 COUNT MEASUREMENTS FROM A COHORT OF HUMAN IMMUNODEFICIENCY VIRUS-INFECTED INDIVIDUALS

3.1. Study population

The Sinikithemba cohort comprises 450 HIV-1 subtype C chronically infected adults enrolled at the McCord Hospital (Durban, South Africa) between August 2003 and 2008. Sociodemographic characteristics, plasma VL and CD4 count measurements were obtained at baseline. The CD4 count and VL were measured every 3 and 6 months, respectively, from enrolment. VLs were determined using the automated CobasAmplicor HIV-1 Monitor test (version 1.5; Roche Diagnostics). CD4 cells were enumerated using the Multitest kit (CD4/CD3/CD8/CD45) on a FACSCalibur flow cytometer (Becton Dickinson). In accordance with the National HIV Treatment Guidelines implemented during the study period, patients were recommended to start ART upon reaching a CD4 count less than 200 cells/mm^3 or WHO stage 3 or 4 symptoms [15]. For the purposes of this particular application, 114 patients who had not returned for a subsequent CD4 count measurement after enrolment or who were not confirmed ARV-naive at study entry were excluded from the analysis, resulting in a cohort of 336 patients. The median CD4 count at enrolment was 357 (interquartile range: 259–509) cells/mm^3 and the mean VL was 4.7 log copies/ml. The overall mean age at enrolment was 33 years and 80% of the patients were women.

3.1.1. Follow-up and censoring. Patients were followed up for a median of 2.48 years (interquartile range: 0.61–4.78 years) and had a median of 8 CD4 count measurements (interquartile range: 3–17 visits). A total of 136 (40%) patients were removed from the study at varying times due to initiation of ART, six of whom were pregnant women. The study ended in 2011, and the follow-up of 124 patients who were still under observation and had not yet initiated ART was terminated. There were 76 (23%) patients who dropped out of the study prematurely. Individuals whose CD4 count measurements were terminated because of study end can be considered to exhibit a dropout pattern that is completely at random [16]. Because ART was recommended at the time of first CD4 count less than 200 cells/mm^3, the reason for removal from the study for these patients can be considered to be at random given that it was based on the observed measurements. The reasons for dropout in the 76 patients was unknown except for five patients who declared relocation as their reason for leaving the study. In this paper, we assume ignorability and conduct likelihood-based analysis. We comment further on explicit modelling of the dropout mechanism in the discussion.

3.1.2. Additional assumptions. Because this is a seroprevalent cohort, the date of the last negative HIV test result is unknown, and hence, the date of seroconversion cannot be estimated without the analysis of additional patients with known dates of infection. For this particular analysis, we have examined two possible timescales: date of first contact as time zero and time on study, which is expressed as the difference between the enrolment date and the date at which the study commenced (1 August 2003). As stated by [11], different timescales in hierarchical models can have a strong impact on the predicted random effects due to the shrinkage effect. We allow a 10-year window prior to enrolment where we consider an individual as having the potential to have experienced the threshold. The rationale for this decision is based on the estimated time from seroconversion to death in ART-naive patients, which was reported to be approximately 10 years in sub-Saharan Africa [17,18].

3.2. Linear mixed model

A variance stabilizing square root transformation was applied to the CD4 count responses. The observed CD4 count trajectories for

F1

eight selected patients is presented in Figure 1. This figure clearly depicts the high degree of between and within individual variability. To explore the relationship between baseline VL and rate of CD4 decline, baseline VL was categorized into approximate tertiles as follows: VL $\leq$ 15 000, 15 000 $<$ VL $\leq$ 100 000 and VL $>$ 100 000 log copies/ml, which represent low, intermediate and high VL, respectively.

There were 92 (27%), 117 (35%) and 127 (38%) patients in the low, intermediate and high VL categories, respectively. We applied the 'general guidelines for model building' recommended in [14], commencing with an elaborate mean structure, which included age, gender, baseline VL and interaction terms as covariates. Although the subject-specific plots and reduced AIC indicated that a quadratic or cubic model may provide a better fit to the observed data than the linear model, these models would result in implausible predicted trajectories outside of the observation period. By comparing nested models using the likelihood ratio test, the inclusion of age and gender did not significantly improve the model fit ($p = 0.150$). The reduced model was of the form:

$$Y_{ij} = \begin{cases} \beta_{0,L} + b_{0i} + \beta_{1,L} t_{ij} + b_{1i} t_{ij} + \varepsilon_{ij} & \text{if 'low' viral load,} \\ \beta_{0,M} + b_{0i} + \beta_{1,M} t_{ij} + b_{1i} t_{ij} + \varepsilon_{ij} & \text{if 'intermediate' viral load,} \\ \beta_{0,H} + b_{0i} + \beta_{1,H} t_{ij} + b_{1i} t_{ij} + \varepsilon_{ij} & \text{if 'high' viral load.} \end{cases}$$

The restricted maximum likelihood estimates with standard errors for each of the timescales are presented in Table I.

Q2

T1

As expected, the model with the time origin as 1 August 2003 resulted in a higher variance of the random intercepts. Through comparison of the AIC and BIC for the two models it is clear that the model with time since enrolment as the timescale provided a better fit to the data. All further analysis was therefore conducted in this timescale, which also facilitates interpretation of the estimated times as times relative to enrolment in the study. We found an overall significant difference in intercepts and slopes between VL categories ($p < 0.0001$ and 0.0053, respectively). Patients with high VL displayed a significantly higher rate of decline in CD4 count than patients with low VL ($p < 0.0001$). More rapid decline in patients with high VL compared with intermediate VL was observed; this result was not statistically significant ($p = 0.115$).

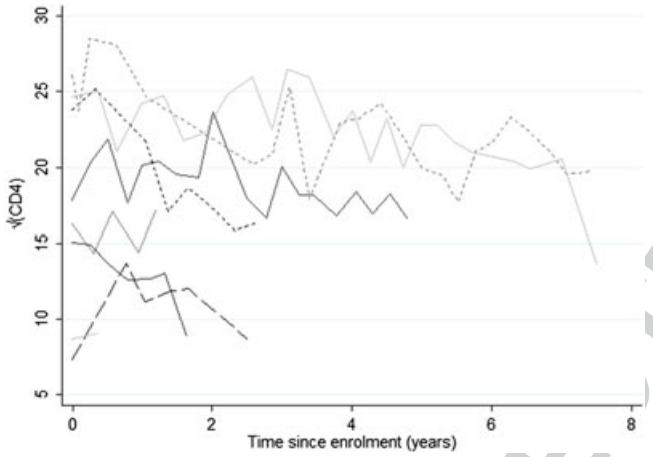


Figure 1. Longitudinal CD4 count measurements for eight subjects on the square root scale.

3.3. Expected time to threshold

As stated in Section 2, we consider an individual to be at risk of obtaining two consecutive values lower than the threshold k up to a maximum of 10 years prior to and post enrolment. The discrete times that fall outside of the observation period were created in accordance with the study design of three monthly visits. The series was truncated at the visit at which the predicted CD4 count $\widehat{Y}_{ij}$ dropped to zero. Similarly, time t_{i1} was defined as the minimum time at which $\widehat{Y}_{ij} < 1500$ cells/mm^3, which is the upper limit of the CD4 count range for HIV infected individuals. We estimated the time to obtain two consecutive measurements less than the threshold values 200 and 350 cells/mm^3, respectively. For ease of presentation, we have chosen to draw attention to the estimation for four specific patients (Figure 2). The estimated

F2

Table I. HIV cohort data. Parameter estimates (standard errors) for the fitted models on each timescale.

Effect	Parameter	Time-enrolment	Time-calendar origin
	Fixed effects estimates (s.e.)		
Intercept	$\beta_{0,L}$	21.2405 (0.4708)	22.0000 (0.5513)
	$\beta_{0,M}$	19.4469 (0.4190)	20.6554 (0.4978)
	$\beta_{0,H}$	16.2821 (0.4057)	17.5021 (0.4909)
Time	$\beta_{1,L}$	−0.5744 (0.1206)	−0.5658 (0.1171)
	$\beta_{1,M}$	−1.0160 (0.1137)	−0.9454 (0.1102)
	$\beta_{1,H}$	−1.3839 (0.1400)	−1.1066 (0.1331)
	Covariance parameter estimates (s.e.)		
var(b_{0i})	d_{11}	19.5555 (1.6080)	25.5456(2.2716)
cov(b_{0i}, b_{1i})	d_{12}	−0.4944 (0.3821)	−2.1611 (0.4703)
var(b_{1i})	d_{22}	0.9941 (0.1421)	0.9438 (0.1303)
Measurement error	σ^2	3.1923 (0.0810)	3.2135 (0.0814)
	Fit statistics		
AIC		17185.3	17225.9
BIC		17200.5	17241.1
−2 REML log-likelihood		17177.3	17217.9

REML, restricted maximum likelihood.

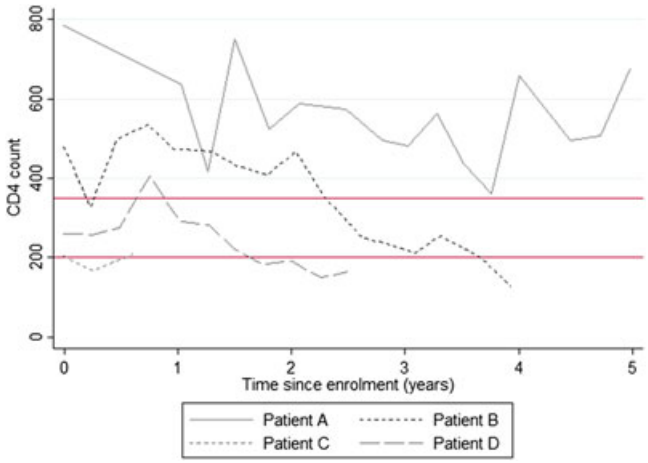


Figure 2. Longitudinal CD4 count measurements with reference at thresholds 200 and 350 cells/mm^3.

probabilities of a single measurement being below the threshold was computed using the estimates from the linear mixed model presented in Section 3.2 and expression (4), which was introduced in Section 2.

Patient A entered the study with a CD4 count substantially above the 200 and 350 cells/mm^3 thresholds and declined at a very slow rate. This is captured by the fitted probabilities where the probability of patient A obtaining a CD4 count less than 200 cells/mm^3 is zero throughout the period considered. The probability of patient A experiencing a CD4 count less than 350 cells/mm^3 increases at 5 years. The estimated time to two consecutive measurements less than the 200 and 350 cells/mm^3 threshold is presented in Table II.

Patient B, who entered the study with a CD4 count above the 350 cells/mm^3 threshold, exhibited a more rapid rate of CD4 count decline than patient A. The estimated time for patient B to reach a threshold of 350 and 200 cells/mm^3 was 2.3 and 4.3 years, respectively. Patients C and D both entered the study with CD4 counts less than 350 but declined at different rates. This is captured by the predicted probabilities in Figure 3. Patient C was estimated to have reached the 200 cells/mm^3 threshold 0.38 years post enrolment, and the 350 threshold 3.26 years prior to enrolment. The confidence intervals for the estimated times for patient C reveal that poorer precision is obtained when analysing individuals with few measurements. Caution should be exercised when interpreting the estimated times for patients who start at a high CD4 count and exhibit a very slow rate of decline. Probabilities of low CD4 count that are zero throughout the period of observation do not pose a problem, but probabilities that increase to greater than zero later in the period can result in estimated times, which are sensitive to the frequency and timing of unobserved measurements which are considered. This is discussed further in Section 3.4.

Several possible analyses using the estimated times can be conducted. We have elected to focus on the estimated probabilities and times themselves as they draw attention to several current issues in the treatment and monitoring of HIV-positive patients. There were 30 individuals who had a zero probability of obtaining a single CD4 count less than 200 cells/mm^3 throughout the period considered. These individuals are referred to as long-term non-progressors. This contributes additional evidence to the proposition that there are individuals who, possibly due to genetics, are able to control the virus. In addition, the estimated times draw attention to a serious public health concern, namely, late presentation for HIV testing. Excluding the individuals who were long-term non-progressors, the percentiles of the estimated times were computed. Fifteen per cent of these patients had already attained a CD4 count less than 200 cells/mm^3 more than 6 months prior to first presentation at the clinic. Hence, patients were choosing to have an HIV test when they were already in the advanced stages of HIV. The ARV treatment guideline in effect during the study recommended treatment initiation at a CD4 count less than 200 cells/mm^3. Therefore, an additional interpretation is that 15% of the patients deviated from the recommended timing of treatment initiation by more than 6 months. After 2011, the treatment initiation cutoff was raised to 350 cells/mm^3. During our study period, we found that 35% of patients had already attained two consecutive CD4 counts less than 350 cells/mm^3 more than 2 years prior to enrolment. It is clear that, unless patients present at the clinic earlier for testing, changing treatment guidelines may not have the desired effect. It would be interesting to examine whether health seeking behaviour has changed over time, by studying individuals who first presented at the clinic after 2011.

3.4. Sensitivity to variation in observation frequency

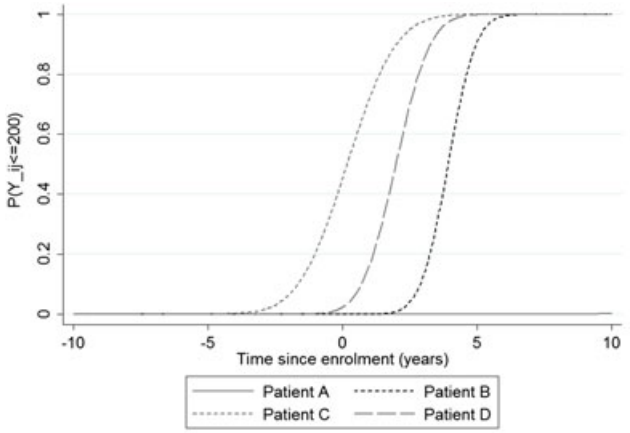
In Section 3.3, we considered an individual to be at risk of experiencing a CD4 count below the thresholds of interest, 10 years prior to and post enrolment, and generated regular three monthly visits for the unobserved period. In this section, we assess the sensitivity of the estimated time to threshold to truncation and the regularity and frequency of visits by simulation. The following scenarios were considered:

Scenario 1. A period of 10 years prior to and post enrolment was considered, and visits outside the observed period occurred at regular three monthly intervals.

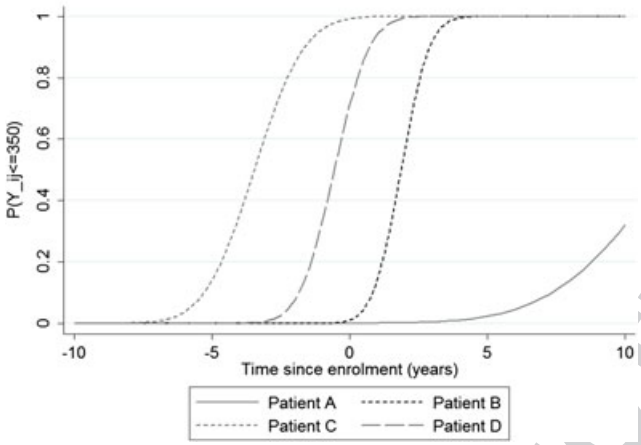
Table II. Estimated time to threshold for patients A, B, C and D.

			$\leq$ 200 cells/mm^3		$\leq$ 350 cells/mm^3	
Patient	VL	Baseline CD4	$\widehat{T}_i$	95% CI	$\widehat{T}_i$	95% CI
A	Low	783	2.92×10^{-5}	$(8.88 \times 10^{-6}, 8.24 \times 10^{-5})$	3.1552	(2.4946, 3.7362)
B	Low	478	4.2858	(4.2343, 4.3843)	2.3046	(2.2887, 2.3169)
C	High	204	0.3758	(0.0267, 0.5319)	−3.2608	(−5.0051, −2.2874)
D	High	261	2.3335	(2.3005, 2.3763)	−0.2043	(−0.4039, −0.0642)

VL, viral load.



a: Probability of single CD4 count less than 200 cells/mm^3



b: Probability of single CD4 count less than 350 cells/mm^3

Figure 3. (a) Probability of single CD4 count less than 200 cells/mm^3 and (b) probability of single CD4 count less than 350 cells/mm^3.

Scenario 2. A period of 5 years prior to and post enrolment was considered. Visits outside the observed period occurred at regular three monthly intervals.

Scenario 3. A period of 10 years prior to and post enrolment was considered, and 10% of visits outside the observation period occurred 1 month later than expected.

Scenario 4. A period of 10 years prior to and post enrolment was considered, and 25% of visits outside the observation period occurred 1 month later than expected.

Scenario 5. A period of 10 years prior to and post enrolment was considered, and 10% of visits outside the observation period were missed.

Scenario 6. A period of 10 years prior to and post enrolment was considered, and 20% of visits outside the observation period were missed.

From Table III, it is clear that the number of visits considered has an impact on the estimated time to threshold for individuals who enter the study with a high CD4 count and exhibit slow decline. Truncation of the series at 5 years results in an estimated time to threshold of 0.005 years, whereas truncation at 10 years results in an estimated value of 3.155 years for patient A. Similarly, the number of visits considered outside the period of observation also has an effect for this type of patient. This is evident from the estimate for patient A under the assumption of 20% of visits being missed. We found that results were far less sensitive for patients experiencing moderate to rapid decline in CD4 count. As expected, in individuals who are likely to reach the threshold of interest during the observed period, the estimated time is robust to truncation or variation in the time points considered. In all patients studied, results were robust to variation in the timing of visits. Hence, the estimated time to threshold is sensitive to the number of time points for specific patients, but not to the actual values of those time points.

4. OTHER AREAS OF APPLICATION

The methodology presented in Section 2 was motivated by the specific problem of estimating the time to CD4 count decline in the presence of persistence criteria, but is not limited to this specific application. Within the HIV context, the methodology is also well suited to prediction of the time to HIV treatment failure, which is defined as the time at which two consecutive VL measurements greater than 400 copies/ml are observed [19]. In settings where less sensitive VL tests are available, the cut-off of interest is 1000 copies/ml [20,21]. More generally, the time to reach two consecutive biomarker measurements lower than or exceeding a threshold is also of interest in the study of non-communicable diseases such as cancer, cardiovascular disease and diabetes. Hence, the methodology presented in Section 2 is well suited to several applications, some of which are briefly discussed in the following.

4.1. Diabetes screening

The glycohemoglobin (HbA1C) test, which measures blood glucose levels over time, is one of the tests which are used to diagnose diabetes. Diabetes is diagnosed when an HbA1C level greater than or equal to 6.5% is observed. As recommended in [22], unless clear symptoms of diabetes exist, diagnosis should only be confirmed when a second measurement of 6.5% or greater is observed. Prediction of the time from screening to diagnosis of diabetes can assist with the development of more efficient monitoring intervals for national screening programmes.

Table III. Estimated time to two consecutive measurements less than 350 cells/mm3 under various scenarios.

Patient	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
A	3.1552	0.0050	3.0734	3.1271	3.0219	2.5941
B	2.3046	2.3046	2.2926	2.2926	2.2926	2.2926
C	−3.2608	−3.2056	−3.2056	−3.2073	−3.2119	−2.9572
D	−0.2043	−0.2043	−0.2030	−0.2097	−0.1432	0.0208

4.2. Recurrence of prostate cancer

According to the European Association of Urology guidelines for prostate cancer, the diagnosis of treatment failure after radical prostatectomy is defined by two consecutive values of prostate-specific antigen greater than 0.20 ng/ml [23]. Several studies have examined risk factors for prostate cancer recurrence using standard approaches, which fail to take into account the inherent trend of prostate-specific antigen and measurement error.

4.3. Abnormal aortic aneurysm screening

In abnormal aortic aneurysm screening studies, surgery is recommended to patients when the diameter of the aneurysm exceeds 55 mm [11]. Because of the high degree of within patient variability in aortic diameter measurements [24], it is of interest to examine the effect of applying persistence criteria in this setting. Estimation of the expected time to threshold for each patient would enable clinicians to identify patients who may be in need of surgery in the near future and target interventions accordingly.

5. CONCLUDING REMARKS

In this paper, we have proposed and applied a novel approach to estimation of the time to attain two consecutive CD4 counts less than two relevant thresholds. This approach takes into account the estimated patient-specific trajectories and measurement error. Through identification of a recursive relationship of the continuation probabilities at each time point, we have displayed that the computation of the expected times is simple, efficient and can be implemented using existing software packages. The method we have proposed can also accommodate complex functions of time, such as quadratic or cubic terms, in contrast to the inverse estimation framework. Additional flexibility in the definition of the event can be achieved by considering the sequences that result in a continuation or 'stop' under the newly defined 'stopping' rule. For example, if the time to the attainment of three consecutive low CD4 counts is of interest, the stopping sequence can be decomposed into a continuation sequence at time point $j-4$ followed by an outcome sequence $\{0, 1, 1, 1\}$.

Sensitivity analysis revealed that the estimated times are sensitive to the number of visits considered and the time at which the series is truncated, for patients who exhibit a very slow decline. For other patients, however, we found that results were less sensitive to the number of visits considered and truncation. Hence, caution should be exercised when interpreting the estimated times for patients who exhibit very slow rates of decline. Another strong assumption that was made for the specific application presented was that visits prior to enrolment and post dropout occurred at regular, equally spaced time points. In the sensitivity analysis conducted, we found that results were robust to deviation from the regular observation times for all patients. The data we have examined is a combination of coarse data due to missing observations and augmented data through the inclusion of random effects. [25] refer to the union of coarsening and augmentation as enrichment. Hence, the prediction of the expected time is dependent on several unverifiable assumptions regarding the random effects structure and missing data mechanism. This raises the importance of conducting sensitivity analysis to gauge the impact of deviations from the assumptions made [16]. Local influence analysis may also be undertaken to determine whether there are specific individuals who have a large impact on the model fit, and hence, on prediction. The primary objective of this paper is to present and apply the proposed methodology to HIV biomarker data. Therefore, local influence analysis and missing data models will be conducted as further work.

To appropriately handle the issue of 'left censored' event times due to the study design, we considered an individual to be at risk of the event of CD4 count decline prior to enrolment. This was achieved by using the date of first contact as the time origin. One may also analyse the data on the timescale of time since seroconversion if additional data from a recently infected cohort is available. This approach would involve imputation of the time of infection via back calculation [26] or inverse estimation. The proposed methodology rests on the assumption that the residual variability is pure measurement error, which may be violated in certain settings. A simulation study that was conducted by [27] to investigate the robustness of fixed effects estimates from a linear mixed model to a misspecified error distribution, revealed that inference is robust when errors are non-Gaussian or heteroscedastic, but may be impaired when errors are correlated. In this latter case, the authors found that the model including a random slope in addition to the random intercept was more robust than the random intercept model. As discussed by [28], extending a linear mixed model proposed by including a more elaborate random effects structure is computationally simpler to implement and can produce practically indistinguishable fits to the data when compared with a model that includes a serial correlation term. In some cases, extensive knowledge of the true underlying process which generates the data, may necessitate the inclusion of serial correlation in the model. Although more computationally intensive, it is possible to relax the assumption of conditional independence in the methodology we have proposed. As a starting point, viewing the process as a first-order Markov chain and applying the rules of conditional probability and Bayes theorem, we are able to express the stopping probability as the product of conditional probabilities. Assuming dependence on the most recently observed value, these conditional probabilities may be expressed as a function of the bivariate and univariate cumulative normal distribution.

Acknowledgement

Geert Molenberghs, Edmund Njeru Njagi and Marc Aerts gratefully acknowledge support from IAP Research Network P7/06 of the Belgian Government (Belgian Science Policy).

REFERENCES

[1] Langford SE, Ananworanich J, Cooper DA. Predictors of disease progression in HIV infection: a review. *AIDS Research and Therapy* 2007; **4**:11.

[2] Burzykowski T, Molenberghs G, Buyse M. *The Evaluation of Surrogate Endpoints*. Springer: New York, 2005.

[3] Mofenson LM, Korelitz J, Meyer WA, Bethel J, Rich K, Pahwa S, Moye J, Nugent R, Read J. The relationship between serum human immunodeficiency virus type 1 (HIV-1) RNA level, CD4 lymphocyte percent, and long-term mortality risk in HIV-1infected children. *The Journal of Infectious Diseases* 1997; **175**:1029–1038.

[4] WHO. Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. *World Health Organization* (2015).

[5] Hughes MD, Stein DS, Gundacker HM, Valentine FT, Phair JP, Volberding PA. Within-subject variation in CD4 lymphocyte count in asymptomatic human immunodeficiency virus infection: implica-

tions for patient monitoring. *Journal of Infectious Diseases* 1994; **169**:28–36.

[6] Malone JL, Simms TE, Gray GC, Wagner KF, Bruge RJ, Burke DS. Sources of variability in repeated t-helper lymphocyte counts from human immunodeficiency virus type 1-infected patients: total lymphocyte count fluctuations and diurnal cycle are important. *Journal of Acquired Immune Deficiency Syndromes* 1990; **3**:144–151.

[7] Crowe S, Hoy J, Mills J. Management of the HIV-infected patient. *Health and Fitness* CUP Archive 1996.

[8] Cardeal da Silva D, Casseb J, Mirzazadeh A, Arruda LB, Rutherford GW. Is the rate of CD4 cell decline changing over time in antiretroviral-naive patients? *AIDS Patient Care and STDs* 2013; **27**:69–70.

[9] Zhang Y, Makvandi-Nejad S, Qin L, Zhao Y, Zhang T, Wang L, Wu H. Interferon-induced transmembrane protein-3 rs12252-C is associated with rapid progression of acute HIV-1 infection in Chinese MSM cohort 2015; **29**:889-894.

[10] Amornkul PN, Karita E, Kamali A, Rida WN, Sanders EJ, Lakhi S, Price MA, Kilembe W, Cormier E, Anzala O, *et al*. Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. *AIDS* 2013; **27**:2775.

[11] Sweeting M, Thompson S. Making predictions from complex longitudinal data, with application to planning monitoring intervals in a national screening programme. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 2012; **11**:569–586.

[12] Mandel M. Estimating disease progression using panel data. *Biostatistics* 2010; **175**:304–316.

[13] Reddy T, Mwambi H, Ndung'u T. Modelling HIV progression using multistate Markov models. *South African Statistical Journal Proceedings: Proceedings of the 53rd Annual Conference of the South African Statistical Association for 2011 (SASA 2011): Congress 1*, 2011; pp. 100–117. Sabinet Online.

[14] Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2009.

[15] SADOH. National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. *Department of Health, Republic of South Africa* 2010.

[16] Molenberghs G, Kenward M. *Missing Data in Clinical Studies*. John Wiley & Sons: New York, 2007.

[17] Van der Paal L, Shafer LA, Todd J, Mayanja BN, Whitworth JA, Grosskurth H. HIV-1 disease progression and mortality before the introduction of highly active antiretroviral therapy in rural Uganda. *AIDS* 2007; **21**:S21—S29.

[18] Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, Mwita W, Suriyanon V, Rangsin R, Nelson KE, Sonnenberg P, *et al*. Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy. *AIDS* 2007; **21**:S55—S63.

[19] DHHS. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. H 2–H 3 2015.

[20] O'Connor J, Smith C, Lampe F, Johnson M, Sabin C, Phillips A. Rate of viral load failure over time in people on ART in the UK Collaborative HIV Cohort (CHIC) study. abstracts of the HIV drug therapy glasgow congress 2014. *Journal of the International AIDS Society* 2014; **17**:19527.

[21] Davies MA, Moultrie H, Eley B, Rabie H, Van Cutsem G, Giddy J, Wood R, Technau K, Keiser O, Egger M, Boulle A. and International Epidemiologic Databases to Evaluate AIDS Southern Africa (IeDEA-SA) Collaboration, Virologic failure and second-line antiretroviral therapy in children in South Africa – the IeDEA Southern Africa collaboration. *Journal of acquired immune deficiency syndromes* 2014; **56**:270–278.

[22] NIH National Institute of Diabetes and Digestive and Kidney diseases (2014). Diagnosis of Diabetes and Prediabetes. 14–4642.

[23] Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, Mason M, Matveev V, Wiegel T, Zattoni F, Mottet N. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative IntentUpdate 2013. *European Urology* 2014; **65**:124–137.

[24] Dapunt OE. The natural history of thoracic aortic aneurysm. *The Journal of Thoracic and Cardiovascular Surgery* 1994; **107**:1323–1332.

[25] Verbeke G, Molenberghs G. Arbitrariness of models for augmented and coarse data, with emphasis on incomplete data and random effects models. *Statistical Modelling* 2010; **10**:391–419.

[26] Taffe P, May M. A joint back calculation model for the imputation of the date of HIV infection in a prevalent cohort. *Statistics in Medicine* 2008; **27**:4835–4853.

[27] Jacqmin-Gadda H, Sibillot S. Proust, C., Molina, J. and Thiébaut, R. *Robustness of the linear mixed model to misspecified error distribution. Computational Statistics & Data Analysis* 2008; **51**:5142–5154.

[28] Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data*. Chapman and Hall/CRC: Boca Raton, 2012.

Appendix

For ease of notation, the index j is suppressed in the equations that follow. We introduce E_j, which denotes the outcome indicator at time point j such that

$$E_j = \begin{cases} 0 & \text{if } Y_j > k \\ 1 & \text{if } Y_j \leq k. \end{cases}$$

Continuation after j visits can be defined in terms of the combination of outcomes observed, such that the event of two consecutive low CD4 count outcomes {1, 1} has not occurred at, or prior to the jth visit. The possible combinations, which lead to continuation after two, three and four visits, respectively, are presented in Table A.1. There are three, five and eight combinations of outcomes that lead to continuation after time points 2, 3 and 4, respectively.

The number of combinations, which result in continuation after each visit, follows a Fibonacci sequence $\{1, 1, 2, 3, 5, 8, 13, \ldots\}$, where each term is defined as the sum of its two predecessors. Specifically, the number of outcome combinations that result in continuation after a sequence of j visits is the $j + 1^{th}$ Fibonacci number, f_{j+1}. Continuation at visit j can be expressed as a function of continuation at visit $j-1$ and $j-2$. A continuation sequence should end in either (A): {0}, which is the union of {0,0} and {1,0}

Table A.1. Possible combinations of outcomes, which result in continuation after two, three and four visits.

E_1	E_2	E_3	E_4
	$j \leq 2$		
1	**0**		
0	**1**		
0	**0**		
	$j \leq 3$		
1	**0**	**1**	
1	**0**	**0**	
0	**1**	**0**	
0	**0**	**1**	
0	**0**	**0**	
	$j \leq 4$		
1	**0**	0	1
0	**1**	0	1
0	**0**	0	1
1	**0**	**1**	0
1	**0**	**0**	0
0	**1**	**0**	0
0	**0**	**1**	0
0	**0**	**0**	0

or (B): {0,1}. This implies that a continuation sequence of length j can be constructed uniquely from (A): a continuation sequence of length $j-1$, followed by {0} and (B): a continuation sequence of length $j-2$, followed by {0, 1}. Letting C_j denote the continuation probability at visit j and assuming that the outcomes at each visit are independent, this recursive relationship can be presented as follows:

$$C_j = C_{j-2} \times P(E_{j-1} = 0) \times P(E_j = 1) + C_{j-1} \times P(E_j = 0).$$

This relationship is illustrated in Table A.1 for $j = 4$. Assuming that the process 'stops' when two consecutive {1, 1} outcomes are observed for the first time, the number of combinations, which result in a 'stop' at sequence of j visits, is f_{j-2}. For a 'stop' to be observed at any $j \geq 3$, the last three outcomes in the sequence are confined to be of the form $\{0, 1, 1\}$. Hence, the stopping probability S_j is

$$S_j = C_{j-3} \times P(E_{j-2} = 0) \times P(E_{j-1} = 1) \times P(E_j = 1).$$

Author Query Form

Journal: Pharmaceutical Statistics

Article: pst_1774

Dear Author,

During the copyediting of your paper, the following queries arose. Please respond to these by annotating your proof with the necessary changes/additions.

- If you intend to annotate your proof electronically, please refer to the E-annotation guidelines.
- If you intend to annotate your proof by means of hard-copy mark-up, please use the standard proofreading marks in annotating corrections. If manually writing corrections on your proof and returning it by fax, do not write too close to the edge of the paper. Please remember that illegible mark-ups may delay publication.

Whether you opt for hard-copy or electronic annotation of your proof, we recommend that you provide additional clarification of answers to queries by entering your answers on the query sheet, in addition to the text mark-up.

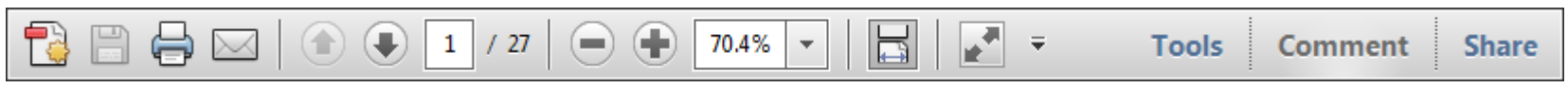
Query No.	Query	Remark
Q1	AUTHOR: Please check if authors and their affiliations are correct.	
Q2	AUTHOR: Is this the correct definition for REML? Please change if this is incorrect.	
Q3	AUTHOR: Please provide city location of the proceeding/conference that was held for Reference 13.	

WILEY

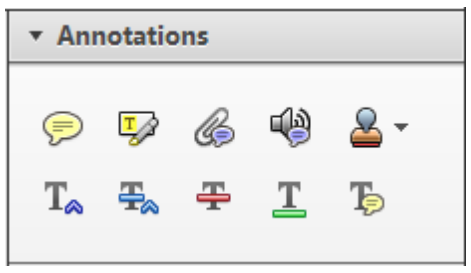
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-Annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 7.0 or above). (Note that this document uses screenshots from Adobe Reader X)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/uk/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:



This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:

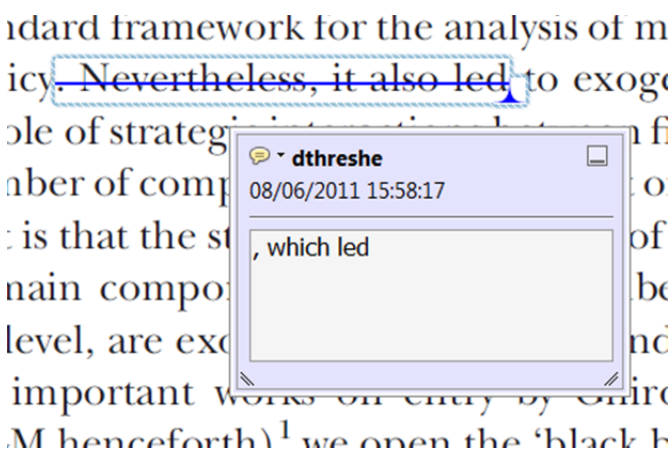


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it

- Highlight a word or sentence.
- Click on the Replace (Ins) icon in the Annotations section.
- Type the replacement text into the blue box that appears.

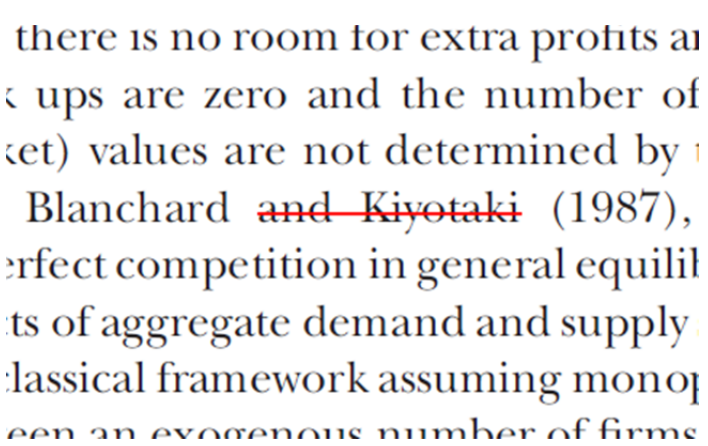


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it

- Highlight a word or sentence.
- Click on the Strikethrough (Del) icon in the Annotations section.

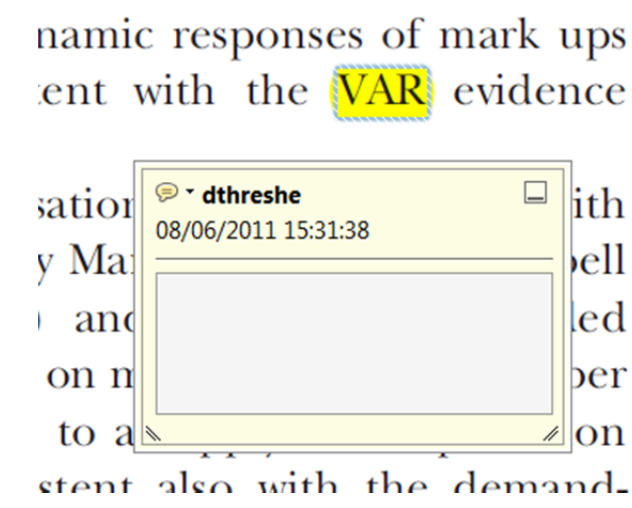


3. Add note to text Tool – for highlighting a section to be changed to bold or italic.

Highlights text in yellow and opens up a text box where comments can be entered.

How to use it

- Highlight the relevant section of text.
- Click on the Add note to text icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.

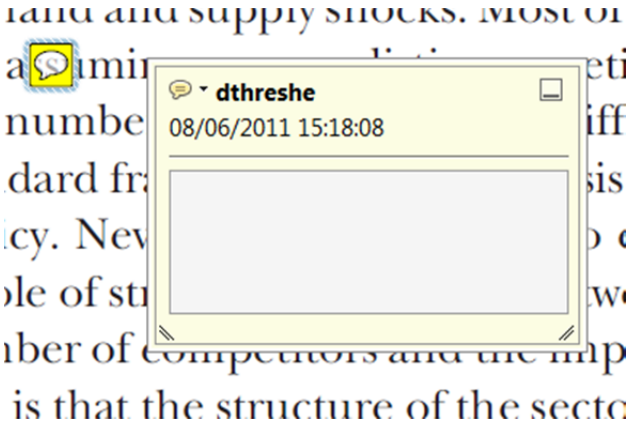


4. Add sticky note Tool – for making notes at specific points in the text.

Marks a point in the proof where a comment needs to be highlighted.

How to use it

- Click on the Add sticky note icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.



WILEY

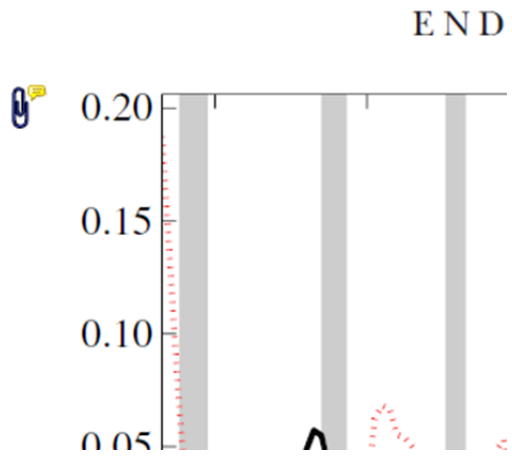
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate pace in the text.

How to use it

- Click on the Attach File icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

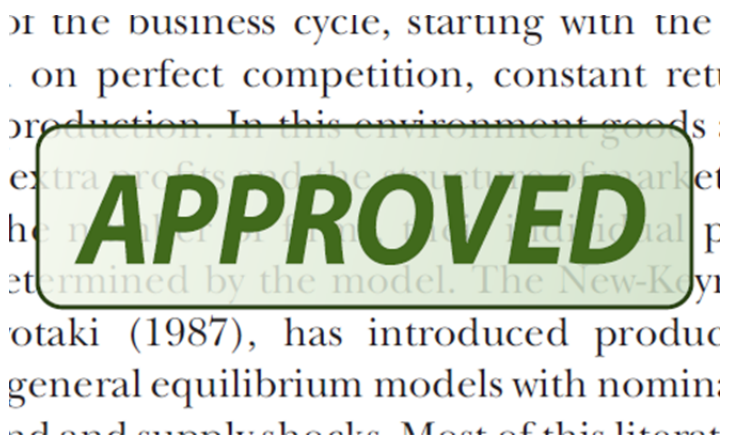


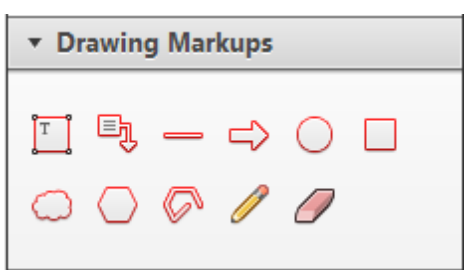
6. Add stamp Tool – for approving a proof if no corrections are required.

Inserts a selected stamp onto an appropriate place in the proof.

How to use it

- Click on the Add stamp icon in the Annotations section.
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears).
- Click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).



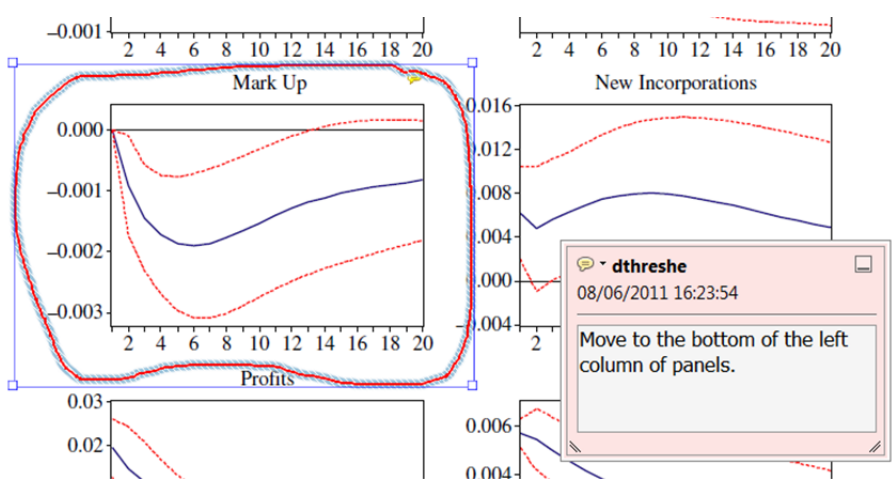


7. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.

Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks..

How to use it

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.



For further information on how to annotate proofs, click on the Help menu to reveal a list of further options: